

## Section 4 XML Overview for CA Business e-file

The FTB uses XML (Extensible Markup Language) Schemas (i.e., “.xsd” files) that specify the overall structure, data elements, and rules of all forms, schedules, and other attachments pertaining to CA Business e-file tax return. This section explains pertinent terms and FTB’s use of XML in our Business e-file Program.

### 4.1 XML Structure

In general, the CA business e-file XML tax return structure is modeled after the IRS MeF 1120 XML e-file structure with the use of XML documents based on paper tax forms. A complete XML tax return is defined by a Return schema. The Return schema contains a ReturnHeader and ReturnData. The ReturnHeader contains general and special processing information about the return that is not stored or repeated within each document attached to ReturnData (i.e., software, filer, preparer, direct deposit information, etc.). The ReturnData contains all of the documents (i.e., tax forms, referenced attachments, etc.) within a return. All of the documents in ReturnData are listed in a specific, and required, sequence (Refer to section 4.6 for how the documents are organized for each return family).

### 4.2 Schemas

Schema (i.e., “.xsd”) files define the structure and format rules for data types, data elements, documents, etc., for each form, schedule, and attachment according to the particular content and context applicable. Schemas have also been defined for general support of a particular form, line, or regulation when the tax forms, instructions and regulations do not provide guidance via a Binary (PDF) or General Dependency attachment.

Within the XML Schema, data elements are the building blocks of an XML document. All elements within the XML schemas have been assigned either a *complex* or *simple* element type. A complex type defines an element that has one or more attributes and/or is the parent to one or more child elements. A simple type defines the data for one element, and may have documentation attributes (i.e., description and line number).

In the Schemas, most data elements for each form, schedule, and the supported documents themselves, have been declared optional. This optional declaration in the Schemas is consistent with the way paper returns are filed (i.e., the taxpayer and return preparer have the responsibility to provide information as specified by the FTB forms, instructions, and regulations).

#### 4.2.1 Tag Names

Each field on a tax form, document, or attachment is identified by XML syntax with a beginning and ending tag written in camel case (see example below). Unlike the current proprietary e-file format for personal income tax returns (as defined in FTB Publication 1346), where field names were assigned a four digit number, XML tag names for business e-file tax returns have been defined using meaningful words and phrases that describe the line, formula, or data being gathered.

For example:

On the CA Form 100 tax form, Line 1 is the first line of the State Adjustments portion of page 1 and is described as *Net income (loss) before state adjustments*.

In an XML return, the tag names and sample data would appear as follows:

```
<NetIncome>2789354</NetIncome>
```

In the XML schema the tag names and annotation describing the element appear as follows:

```
<xsd:element name="NetIncome" type="irs:USAmountType" minOccurs="0">
  <xsd:annotation>
    <xsd:documentation>
      <Description>Net income or loss before state adjustments.</Description>
      <LineNumber>Line 1</LineNumber>
    </xsd:documentation>
  </xsd:annotation>
</xsd:element>
```

**Note:** Usage of the XML annotation construct to document each element's description and line or part reference appears throughout all CA Business e-file XML Schemas. Please refer to these annotations to insure proper usage of the XML as it relates to the tax return forms.

#### 4.2.2 Attributes

Attributes provide additional information or describe a feature of low level data elements, groups of elements (i.e., parts and sections within tax forms), and document elements. Attribute names begin in lower case, and are completed in camel case. Every return document, form, schedule, or supporting attachment document, has been defined with the following attributes in its root element (also, see Sample XML Data below):

- **documentId:** Required attribute that uniquely identifies the document within the context of the whole return. Tax return preparer's software is responsible for generating a unique id of idType defined in the IRS' efileTypes.xsd file for each of the return documents.
- **documentName:** Optional attribute of string type with a FIXED value. If it is used, it must have a value equal to the name of the form, schedule or a supporting document it represents as given in the Schema.
- **softwareId:** For California purposes, this is your CTP ID issued by our substitute forms program.
- **softwareVersion:** The version of your software that you provide.

#### Sample XML Data

```
<CAForm100ScheduleH documentName="CAForm100ScheduleH" documentId="A2"
softwareId="12345678" softwareVersion="S2006v3.2">
```

#### 4.2.3 Repeating Groups

The repeating group is a data element or a group of data elements that can repeat the number of times specified in the *maxOccurs* attribute. Unlike paper tax forms, most e-file XML repeating groups are set to allow an unlimited amount of information to be provided.

For example, Schedule D, Part 1, Line 1, Short-Term Capital Gains and Losses, allows two entries on the paper form. The corresponding Business e-file Schema allows an unlimited number of entries.

The following example illustrates a repeating group schema where complex types define a content model for a repeating group of data. The attribute that makes the group repeat appears highlighted. Also appearing below is sample XML data from this structure:

```
<xsd:complexType name="CA-RoyaltiesScheduleType">
  <xsd:annotation>
    <xsd:documentation>Content model for CA Non-Business Royalties Schedule</xsd:documentation>
  </xsd:annotation>
  <xsd:sequence>
    <xsd:element name="RoyaltyItem" type="CA-RoyaltiesType" minOccurs="0"
      maxOccurs="unbounded"/>
  </xsd:sequence>
</xsd:complexType>

<xsd:complexType name="CA-RoyaltiesType">
  <xsd:sequence>
    <xsd:element name="Explanation" type="irs:ShortExplanationType" minOccurs="0"/>
    <xsd:element name="Amount" type="irs:USAmountType" minOccurs="0"/>
  </xsd:sequence>
</xsd:complexType>
```

#### Sample XML Data

```
<RoyaltyItem>
  <Explanation>Sample Explanation 1</Explanation>
  <Amount>1000</Amount>
</RoyaltyItem>
<RoyaltyItem>
  <Explanation>Sample Explanation 2</Explanation>
  <Amount>2000</Amount>
</RoyaltyItem>
<RoyaltyItem>
  <Explanation>Sample Explanation 3</Explanation>
  <Amount>3000</Amount>
</RoyaltyItem>
```

#### 4.2.4 Choice Construct

The choice construct within an XML Schema batches a data element or group of data elements between choice tags and only allows one of the data elements or groups to be provided in an instance. The following depicts a choice between a USAddress or ForeignAddress:

```
<xsd:choice>
  <xsd:element name="USAddress" type="irs:USAddressType"/>
  <xsd:element name="ForeignAddress" type="irs:ForeignAddressType"/>
</xsd:choice>
```

#### 4.2.5 Union Construct

The union construct defines a simple data type as a collection (union) of more than one data type. The following depicts use of the union construct to define the CA-CorporationPaymentType:

```
<xsd:simpleType name="CA-CorporationPaymentType">
  <xsd:annotation>
    <xsd:documentation>Defines CA corporation payment types</xsd:documentation>
  </xsd:annotation>
  <xsd:union memberTypes="CA-ReturnPaymentType CA-EstimatePaymentType"/>
</xsd:simpleType>
```

Both of the member types above consist of enumerated values describing the allowable payment types. See illustration below:

```
<xsd:simpleType name="CA-ReturnPaymentType">
  <xsd:annotation>
    <xsd:documentation>Defines a CA return payment type</xsd:documentation>
  </xsd:annotation>
  <xsd:restriction base="irs:StringType">
    <xsd:enumeration value="Return"/>
  </xsd:restriction>
</xsd:simpleType>
<xsd:simpleType name="CA-EstimatePaymentType">
  <xsd:annotation>
    <xsd:documentation>Defines CA estimate payment periods or types</xsd:documentation>
  </xsd:annotation>
  <xsd:restriction base="irs:StringType">
    <xsd:enumeration value="1stQtrEstimate"/>
    <xsd:enumeration value="2ndQtrEstimate"/>
    <xsd:enumeration value="3rdQtrEstimate"/>
    <xsd:enumeration value="4thQtrEstimate"/>
  </xsd:restriction>
</xsd:simpleType>
```

The resulting allowable values for the union of the CA-ReturnPaymentType and CA-EstimatePaymentType into the CA-CorporationPaymentType are: Return, 1stQtrEstimate, 2ndQtrEstimate, 3rdQtrEstimate, and 4thQtrEstimate.

#### 4.2.6 e-file Types

e-file types are defined global data types that are used in more than one, and for some types, all schemas for element definitions. Because of this, they are often referred to as “base” data types. Following are the 5 California base e-file type schema files and 1 IRS base e-file type schema file that provide the foundation of data type definitions for California Business e-file XML tax returns:

- **CA-eFileTypes** (supports all California BEeF schemas)
- **CA-PartnershipTypes** (supports CA Form 565 and 568 families)
- **CA-CorporationTypes** (supports CA Form 100 and 100S families)
- **CA-InterfaceTypes** (supports CA Transmission schemas)
- **CA-EDATypes** (supports CA Economic Development Area (EDA) Forms 3805Z, 3806, 3807 & 3809)
- **efileTypes** (supports all California and Federal XML tax return schemas)

The base schema files include a wide variety of types, such as:

- Short simple types (for example: **CA-SOSNType**: Type for California Secretary of State file number) that define the restrictions (e.g., facets, patterns and enumerations) of acceptable data
- Large complex types composed of many elements with intricate structures (for example: **CA-DistributiveltemsType**: Content model for Forms 565 and 568 - Schedules K and K-1 - Partners or members shares of income, deductions, credits, etc.).

Global or base data type definitions also appear in document or form level schemas. These definitions only apply to elements within the structure of the document defined by that schema. For example, within the CAForm100ScheduleH schema the CA-IntercompanyDividendType defines elements for repeating rows in part 1 of the CAForm100ScheduleH.

Generally, FTB has made a commitment to utilize the IRS efileTypes whenever possible. However, we have established our own CA Types when no comparable federal element or structure exists, and/or when CA law, forms, instructions, etc., differ from federal requirements.

#### 4.2.7 Re-Use of Complex Types

Throughout the CA Business e-file XML schemas where information requested on various portions of the tax forms are identical, or substantially similar, we often have created only one structure to satisfy multiple requirements (please see [partial](#) listing of examples below). This design avoids duplication and results in less physical schema structure. To accommodate this design where the structure addresses more than one tax form, but where the line number references are different, we have included additional guidance in the annotated descriptions and line number documentation. Please review this documentation for detailed information on which elements apply to particular tax forms and specific lines.

##### CA-CorporationTypes Re-Used Structures

- **CA-BalSheetAssetType** (Content model for Form 100 and Form 100S Schedule L - Assets Balance Sheet.)
- **CA-BalSheetLiabStockEqType** (Content model for Form 100 and Form 100S Schedule L - Liabilities and Stockholder Equity Balance Sheet.)
- **CA-ProrataShareltemsType** (Content model for Forms 100S - Schedules K and K-1 - S Corporation Shareholder's Pro-rata share of income, deductions, credits, etc.)
- **CAForm100SeriesScheduleJType** (Content model for Form 100 and Form 100S Schedule J - Add-On Taxes and Recapture of Tax Credits)
- **CAForm100SeriesScheduleLType** (Content model for Form 100 and Form 100S Schedule L - Balance Sheet)
- **CAForm100SeriesScheduleVType** (Content model for Form 100 and Form 100S Schedule V - Cost of Goods Sold)

## CA-PartnershipTypes Re-Used Structures

- **CA-DistributiveItemsType** (Content model for Forms 565 and 568 - Schedules K and K-1 - Partners or members shares of income, deductions, credits, etc.)
- **CA-PartnershipLLCEntityInformationType** (Content model for CA Forms 565 and 568 - Schedule K-1 - Partnership or LLC entity information and questions)
- **CA-PartnershipLLCIncomeType** (Content model for CA Form 565 and Form 568 Schedule B - Income Lines)
- **CA-PartnershipLLCDeductionsType** (Content model for CA Form 565 and Form 568 Schedule B - Deductions Lines)
- **CA-Form565Form568ScheduleAType** (Content model for CA Forms 565 and 568 - Schedule A - Cost of Goods Sold)
- **CA-Form565Form568ScheduleLType** (Content model for CA Forms 565 and 568 - Schedule L - Balance Sheet)
- **CA-Form565Form568ScheduleM-1Type** (Content model for CA Forms 565 and 568 - Schedule M-1 Reconciliation of Income Per Books/Return)
- **CA-Form565Form568ScheduleM-2Type** (Content model for CA Forms 565 and 568 - Schedule M-2 Partners' Capital Accounts)
- **CA-Form565Form568ScheduleDType** (Content model for CA Forms 565 and 568 - Schedule D - Capital Gain or Loss)

### 4.2.8 Identity Constraints

Like the IRS, CA Business e-file XML schemas employ identity constraints that require a document's documentId to be unique across a return. The constraint for each return can be found at the highest, or outermost element of each return family named *CA-Return*. The following depicts the definition of the CAReturn100's identity constraint:

```
<xsd:unique name="documentId">  
  <xsd:selector xpath="ftb:CA-ReturnData/ftb:*|ftb:CA-ReturnData/irs:*/>  
  <xsd:field xpath="@documentId"/>  
</xsd:unique>
```

### 4.3 Attachments to CA Business e-file Returns

Taxpayers are instructed to attach supporting information to the tax return at specific points of reference. This includes tax forms, statements, elections, notices, schedules, and other types of miscellaneous information that may be required by the FTB Forms, instructions, regulations, or publications. These may be submitted within the CA Business e-file XML return as XML documents or as binary files (i.e., "pdf" files).

#### 4.3.1 XML Attachments

XML attachments may only be attached where specific references are noted within the XML schema. Attachments that are prepared and submitted in this fashion are identified by the referenceDocumentName attribute associated to the element or document they apply to. Each of these names refers to a supporting schema located in the default California or Federal "irs:" namespace. XML documents referenced from within an XML return instance should only be the documents identified within the referenceDocumentName.

For attachments not specifically pre-identified, pointed to or provided for within the schemas, we have provided the CA-GeneralDependency and CA-BinaryAttachment (i.e., ".pdf") schemas. The CA-General Dependency schema includes elements that provide the:

- **FormLineOrInstructionReference** – the form and/or line number to which the attachment belongs.
- **RegulationReference** – the regulation and/or publication reference.
- **Description** – provide a meaningful title or explanation for the attachment.
- **AttachmentInformation** – additional information regarding the attachment.

### 4.3.2 Binary Attachments

The CA Business e-File tax return may also include non-XML documents, known as *binary attachments*, submitted in “.PDF” format. These attachments are part of the Submission Archive file, and only information about them is located inside the XML return data. They allow taxpayers to provide information and documentation that is not provided for in a defined XML Schema. This includes information such as the form FTB 8453-C, *California e-file Return Authorization for Corporations*, and other miscellaneous, schedules, explanations and other information that may be provided in support of the return.

Unlike the IRS, we do not identify and locate the binary attachment within the Schema by use of a ReferenceAttachmentID. We only require that for every binary attachment there is an accompanying CA-BinaryAttachment XML document. The CA-Binary Attachment includes elements that provide the:

- **FormLineOrInstructionReference** – the form and/or line number to which the attachment belongs.
- **RegulationReference** – the regulation and/or publication reference.
- **Description** – provides a meaningful title or explanation for the attachment.
- **AttachmentLocation** – Information indicating the file path name of the binary attachment (PDF).

### 4.4 Namespace

Namespaces in XML Schemas allow all the vocabulary of type definitions that occur within or under a particular namespace to remain separate and unaffected from definitions in other namespaces. XML vocabularies can co-exist within a schema if they are assigned different monikers or prefixes. We utilize three different Namespaces in CA Business e-file:

- **The California namespace** (<http://www.ftb.ca.gov/efile>) that does not utilize a moniker or prefix, and is the default namespace for all schemas that are utilized in the California XML return.
- **The IRS’ namespace** (<http://www.irs.gov/efile>) which has the moniker/prefix “irs:” Note: within the IRS XML return submission, this namespace is the default namespace and does not have or utilize a moniker or prefix.
- **The W3C namespace** (<http://www.w3.org/2001/XMLSchema>), which utilizes the moniker/prefix “xsd:”. This namespace contains base definitions that govern and define the use and syntax of XML.

### 4.5 Import Statement

The Import statement allows use of the schemas from a different namespace within the default namespace. In order to use of the federal schemas (e.g., base data types and attachment schemas) in their unaltered state within the California XML return, we use the XML Import statement.

### 4.6 Return Data Organization

The ReturnData schema files for the CA Form 565, 568, and 100S return families have been organized so that they are easy to locate and appear in the order of their relative importance/significance. First, the documents within appear by the general context to which they apply (see CA-ReturnData565 example below). Within each contextual area, documents are then ordered in alphabetical and numerical sequence.

Required Forms

- CAForm565

CA Forms

- CAForm565Form3885P CAForm565ScheduleD CAForm565ScheduleK-1 CAForm592B...

IRS Forms

- irs:IRS982 irs:IRS4562 irs:IRS4797 irs:IRS6781 irs:IRS8082 irs:IRS8275...

CA Referenced Attachments

- CA-AdditionalSection263ACostsSchedule CA-AllocableMiscIncomeLossSchedule CA-AllocableRoyaltiesSchedule...

- [irs:IRS970](#) [irs:IRS3115](#) [irs:IRS4684](#) [irs:IRS8271...](#)

## **4.7 Schema Validation**

One of the most significant benefits of using XML and schemas for e-file tax returns is that the XML instance documents (i.e., returns) can be validated against the schemas that define the structure and data types, prior to submitting the return for further processing. This provides the advantage of checking for errors as early as possible. We strongly encourage all CA Business XML e-file tax returns to be validated against the current production schemas (as noted on our Website) prior to transmission to the FTB. Schema validation errors are the biggest reason we reject business e-file returns.

## **4.8 Business Rule Validation**

Not all requirements for CA Business e-file returns can be satisfied by schema definitions and schema validation. If an XML return passes schema validation, it is then processed against our business rules. Business rules enforce the business requirements as prescribed by our forms, instructions, laws, and regulations. Returns failing business rule validation are rejected and are considered not filed. Please refer to Section 6.7.2 for more information about our business rules and business rule validation.